

Ontology based Semantic Querying of the Web using Protégé

Meena Unni

Department of Computer Science
Karpagam University
Coimbatore, 641 024, India

K. Baskaran

Associate Professor, Department of Computer Science
Government College of Technology
Coimbatore 641 015

ABSTRACT

Currently, the Web is primarily composed of documents written in HTML. Information on the web is designed for human consumption. Retrieval of information on web is mostly restricted to keyword searches. This limitation may be overcome by semantic web which is an intelligent and meaningful web proposed by Sir Tim Berner Lee. Ontology is the structural framework in semantic web and plays a crucial role in information exchange, knowledge reuse and provides reasoning capabilities for human and machines. Ontology is a set of concepts and the relationships that exist between those concepts. Components of ontology are individuals, classes and properties.

Protégé along with DL Query helps to create a consistent ontology and conduct reasoning based on the concepts and relationships from the domain knowledge. This paper explains the use of Protégé 4.0 to create OWL ontology for Vitamin A and its effects on humans. This ontology represents information about Vitamin A and its effects on humans, Personal factors influencing such effects, and dietary sources which are responsible for these effects. It also infers groups of people who are vulnerable to vitamin A inadequacy. This paper also illustrates the usefulness of ontology in the creation of “web of data”

General Terms

Protégé, Semantic Query, ontology editor

Keywords

Protégé, OWLViz, DL Query, Semantic web, Ontology

1. INTRODUCTION

The World Wide Web as we know today is a huge repository of data. It is based on HTML and different documents are interlinked using hypertext. As the information available on the web increases, it becomes increasingly difficult to provide effective search access for information retrieval. Search engines like Google, yahoo are used to search for information on the web. Search engines do their search and retrieval using keywords. A query on human would not return a document that used the word ‘people’. Such problems can be overcome using semantic web.

Semantic web presents users with an ability to work on shared meaningful knowledge representations. It makes the meaning clear so that a computer can understand it or at least utilize it. Fundamental principle of semantic web is the creation and use of semantic metadata. Ontology is the heart of semantic web. Ontology is an explicit and formal

specification of a conceptualization of a domain of interest [5].

They are used as a unified framework or shared understanding of communication between people and systems [3]. Ontology can be used to share information between people and machines. Different websites may contain disparate details of vitamin A. All this information can be combined using ontology. This combined information can be used to answer user queries. This ontology can be reused in the future. If the domain is large, a large ontology can be built by integrating existing ontologies each describing part of large domain. It is also possible to edit, modify or update the current ontology.

The purpose of this paper is to build ontology for the domain of vitamin A. Protégé is used to create ontology and Fact++ as the reasoning tool. This ontology includes effects on human, personal factors, dietary sources. Facts are used to check the consistency and to infer knowledge.

In the following section, after a brief introduction, taxonomy and general structure of the ontology is discussed, highlighting classes and properties. The ontology is implemented using Protégé 4.0.1 and DL Queries is used to infer knowledge. Finally conclusion is given and possible ways of using ontology is suggested.

2. TAXONOMY

Vitamin A is one of the most essential vitamins required for the well functioning of human body.

Vitamin A deficiency is a major nutritional concern in underdeveloped and developing countries. India is listed as one of the top ranking countries by the world health organization with high degree of vitamin A deficiency in the population [4]. The major cause of deficiency is a diet which provides little or no amount of the nutrient. It is also the leading cause of preventable childhood blindness. Vitamin A is a vital nutrient needed for the normal functioning of vision, growth and development, reproduction and immunity [8].

Human body cannot synthesize vitamin A and hence a diet rich in the nutrient has to be consumed. When dietary intake is continually low, vitamin A will be inadequate to support vision and cellular development resulting in weakened tissue function. Deficient intake of vitamin A during pregnancy or lactation increases the risk of vitamin A deficiency disorders.

Vitamin A is stored in the liver. Infections, such as measles, diarrhea or respiratory diseases deplete body stores of vitamin A. Excess intake of vitamin A over long period of

time can finally go over the liver's storage capacity and produce adverse effects.

Creating taxonomy for Vitamin A will augment the common man's knowledge in the area. The domain knowledge can be used by the nutritionist.

It is difficult to build an intelligible ontology in a single attempt. It requires several iteration and collaboration among individuals who are expertise in specific area of the domain. We tried to create ontology of vitamin A, its types and effects in minimum requirement basis.

2.1. Vitamin A

There are two forms of Vitamin A: Preformed Vitamin A and Provitamin A Carotenoids. Foods from Animal sources which include dairy products, fish and meat contain preformed vitamin A. Beta-carotene is the most important provitamin A and is found in foods from plant sources.

2.2. DietarySource

Food from animal sources and vegetable sources contain vitamin A. The richest food source of preformed vitamin A is meat, especially liver. Egg, milk, butter and cheese too contains significant amount of the vitamin. The richest source of provitamin A carotenoids are found in carrot, sweet potato, green leafy vegetables, yellow/orange vegetables and fruits.

2.3. EffectsonHuman

Vitamin A plays a central role in a variety of functions throughout the body like vision, immune function, growth and development, bone metabolism, skin and cellular health.

Inadequate amount of vitamin A results in vitamin A deficiency disorders. Deficiency will lead to eye diseases like impaired vision or skin diseases like eczema or other types of ailments like anemia. Consuming large doses of vitamin A over a long period of time will result in ailments like blurry vision or headache.

2.4. Person

Recommended Dietary Allowance (RDA) is the amount of vitamin A measured generally in micro grams that is sufficient to meet the nutrient requirement of a healthy individual. This amount varies depending on age and sex. Tolerable Upper Intake Level (UL) is the maximum daily amount of vitamin A that will not cause any adverse health effects in a healthy individual [3].

RAD and UL varies depending on age and sex. Due to time restriction, instead of specifying RDA and UL for all age categories, we have reduced the number of subclasses to represent infant, child, male and women. Infant class specifies infants between the age of 1 and 3. Child has two subclasses: between 4 and 8 and between 9 and 13. Female class is further divided into Women, lactating or Pregnant.

2.4.1. Inferred Class

VulnerableGroupofVitaminAInadequacy is an inferred class of individuals who are likely to have inadequate intake of Vitamin A. Pregnant and lactating women who require extra vitamin A for their own metabolism, fetal growth and tissue maintenance, are likely to intake insufficient amount of the nutrient [3]. People suffering from certain diseases like asthma, chronic renal failure and osteoporosis increase risk

of deficiency due to difficulty in absorbing. Supplements are required for this group of individual.

2.5. Interaction

Vitamin E interacts positively with vitamin A. Adequate amount of Vitamin E protects status of vitamin A in the body. On the other hand certain infection like measles confronts the status of vitamin A. Heavy intake of alcohol impairs liver thus diminishing the storage of vitamin A. Besides protein, zinc and iron deficiency interferes with vitamin A metabolism.

3. IMPLEMENTATION

With tremendous success of semantic web, many ontologies are available on the web. It is possible to use an existing ontology and add more knowledge to it.

There is no one correct way to model domain— there are always viable alternatives. The best solution almost always depends on the application that you have in mind and the extensions that you anticipate [1].

Ontologies specify concepts and relationship between concepts. These concepts and the relationships between them are usually implemented as classes, relations, properties, attributes, and values (of the properties/ attributes) [5]. The methodology used to develop this ontology is methontology[9]. According to Noy and McGuinness [1], the following phases are to be followed to develop ontology.

3.1. Determine the domain and scope

Development of ontology should start by defining its scope and domain. Following competency questions have been considered to determine the scope.

What are the effects of deficiency of Vitamin A?

What is the recommended dietary allowance of vitamin A?

Which foods are rich in Vitamin A?

What are the different types of Vitamin A?

Since vitamin A is one of the most important nutrients required for the accurate functioning of the body, we have taken vitamin A as the domain. Intake of insufficient amount of vitamin A results in night blindness which could lead to complete blindness. It is surprising to note that many Indians especially from lower income groups suffer from vitamin A deficiency [4]. In fact all they require is vitamin A supplement. Overdose of vitamin A too is toxic. It is advisable not to exceed the nutrient more than the tolerable upper level. This ontology could be used by all to retrieve relevant details of the nutrient. Knowledge base of the ontology can be extended by nutritionist.

3.2. Consider reuse

There are ontologies available from third party that provides a starting point. We carried out some research for any existing ontologies on vitamin A, but could not locate any. Hence we developed from scratch.

3.3. Enumerate Key terms

It is useful to list all the terms that is likely to appear in the ontology. Classes, relationship between them and instances can be listed. For our domain, important terms include types, animal food, plant food, deficiency, overdose, personal factor etc.

3.4. Define taxonomy

Once the key terms are identified, they have to be arranged in class-subclass relationship. Class hierarchy can be developed using top-down, bottom-up or combinational approach. We have used top down developmental approach.

3.5. Define the properties of the classes

A property defines relationship between object. If the relationship is between two objects, it is an object property. Data property defines relationship between an individual and data values. Enhances, Isenhancedby, Causes are examples of object properties. Enhances defines relationship between dietarysource and withinlimits. Whereas hasquantity, hasvitainmcg are examples of data property. Hasvitainmcg gives the quantity of vitamin A in an instance say plant food carrot.

3.5. Define the facets of a class

A facet of a property explains the value type, allowed values, the cardinality and other features of the values the property can take [1]. For example consider the property hasvitainmcg. The value type of this is 'integer' and cardinality is one.

3.6. Define instances

This is the last step in creating ontology. Instances can be manually typed or retrieved from another source like database.

4. IMPLEMENTATION IN PROTÉGÉ

4.1. Classes and Sub Classes

Classes are the building blocks of ontology. Top level ontology classes are as shown.

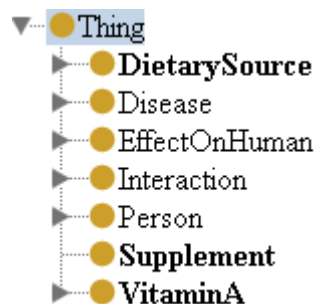


Fig 1- Main Classes in Protégé

Classes can have subclasses which show next level taxonomy. The following figure shows subclasses of dietary source. Two subclasses of Dietary source are animalsource and plantsource. Dairy, egg, fish, meat are the subclasses of Animal source. Fruit, vegetable, lentil and wholegrain are the subclasses of plantsource.

Vitamin A requirement differs according to age and sex. To take into account all the age categories will be time consuming and cumbersome effort, hence we have broadly

sub classed them as infant, child, female and male. There is massive difference between children ages 4 to 13; Hence we have sub classes them as between4-8 and between9-13. In female subclass, pregnant and lactating women requirement of vitamin A is considerably higher than those of normal women. Hence female is sub classed into Women, Lactating and Pregnant.

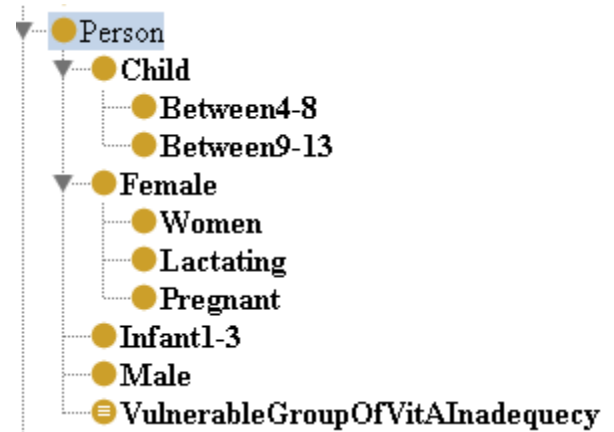


Fig 2 – Class Person and its subclasses

The main reason for developing this ontology is to find the effect of vitamin A on human i.e. to know the consequences of deficiency, overdose and if taken within limits. All these classes have sub classes. For example deficiency is divided into eyedisease, skindisease and otherDisease.

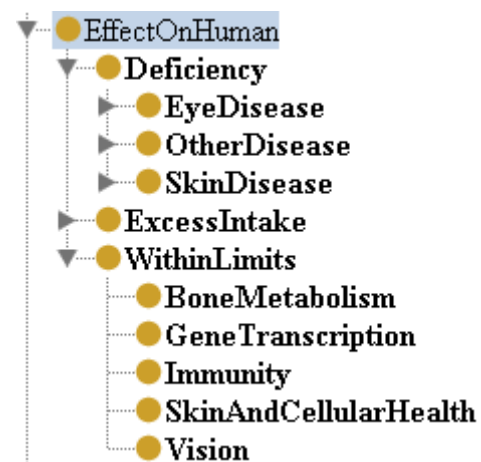


Fig 3- Subclasses for EffectOnHuman

4.2. Properties, facets and instances:

In the vitamin A ontology, object and data properties have been defined. Instances have been created. This will also help us to test the inferred knowledge

4.3. Class hierarchy in OWL Viz

Here we display some visualization results of the ontology. Protégé provides a plug in called OWL Viz. This can be used to view asserted and inferred class hierarchy.

Figure 4 shows the asserted class hierarchy of forms of vitamin A.

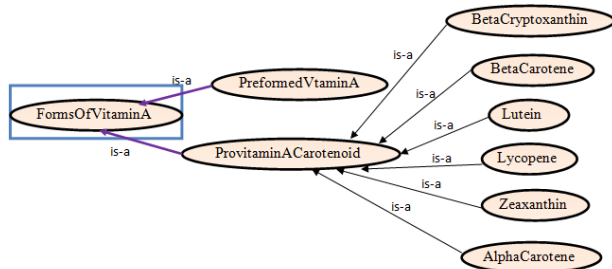


Fig 4 – OWLViz diagram for forms of VitaminA

Figure 5 shows the asserted class hierarchy of dietary source of vitamin A.

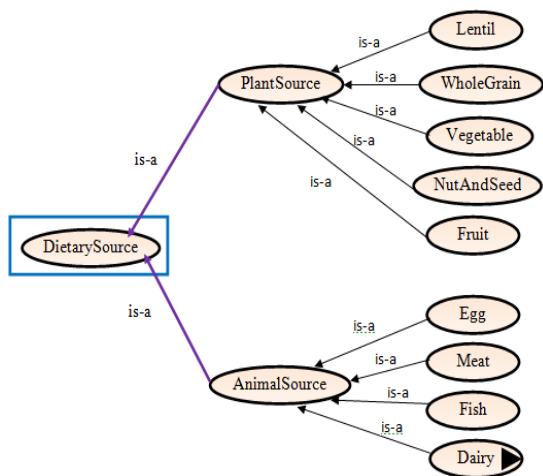


Fig 5- OWLViz class diagram for DietarySource

5. DESCRIPTION LOGIC QUERY

Till now we had explained the design and implementation of our ontology. This part focuses on retrieving knowledge from the ontology. DL Query tab along with the reasoner Fact++ is used to retrieve all instances or related classes. Below we give some examples of queries answered by the ontology.

Figure.6 illustrates a query for deficiencies that can be prevented with proper diet. It answers with the list of deficiencies using the facts from the ontology.

What are the deficiencies that can be prevented by some diet?

Query (class expression)
 Deficiency and IsPreventedBy some DietarySource

Query results

- Diarrhea
- EarInfection
- Eczema
- EyeDisease
- Glucoma
- Hyperkeratosis
- Ichthyosis
- ImpairedVision
- Influenze
- KeratosisFollicularis
- LowBonedensity

Fig 6 – DL query Example 1

Figure.7 illustrates people who are at risk of having Vitamin A inadequacy.

Which Groups are at high risk of Vitamin A inadequacy?

Query (class expression)
 VulnerableGroupOfVitAInadequacy

Query results

Sub classes (2)

- Lactating
- Pregnant

Instances (4)

- ◆ Person7
- ◆ Person3
- ◆ Person5
- ◆ Person1

Fig 7 – DL Query Example 2

Figure.8 illustrates a query for any plant food that has more than 500 mcg of vitamin A. It returns only two individuals as only these two individuals are entered in the ontology.

The screenshot shows a query interface with the following elements:

- Query (class expression):** A text box containing the query: `PlantSource and HasVitAInMcg some int[>500]`
- Buttons:** "Execute" and "Add to ontology"
- Query results:** A section titled "Instances (2)" containing two entries:
 - ◆ SweetPotato
 - ◆ Carrot

Fig 8 DL Query Example 3

Figure.9 illustrates a query for negative interaction of vitamin A

The screenshot shows a query interface with the following elements:

- Query (class expression):** A text box containing the query: `Interaction and DisAgreesWith some VitaminA`
- Buttons:** "Execute" and "Add to ontology"
- Query results:** A section titled "Descendant classes (6)" containing six entries:
 - Alcohol
 - IronDeficiency
 - Measle
 - NegativeInteraction
 - ProteinDeficiency
 - ZincDeficiency

Fig 9 DL Query Example 3

6. CONCLUSION

This paper focusses on describing the information process for both human and machine understanding. This work aims to fulfill the semantic web objectives by describing the information intelligently. Ontology is one of the key concepts towards this vision.

Users spend considerable amount of their time browsing and reading to find out how information is related to one another and where it falls in the overall structure of the domain. Interrelationship between information can put isolated information into a meaningful context. This can be facilitated through ontologies. Different websites may contain different information about vitamin A. Through ontology these information can be aggregated and can be published. The users can use aggregated information to answer queries. The developed ontology can be reused in the future for other users. To build ontology for a large domain, say nutrition, existing ontology describing the

portions of the large domain can be integrated. Ontology can be updated too if required.

This paper describes how we built vitamin A ontology using protégé. We have used N.Noy, D.L.McGuinness steps for creating the ontology. Creating a good ontology for a specific domain relies on adopting a suitable methodology and following the guidelines while modeling and implementation process. Nevertheless, implementation of the ontology is subject to functionalities of the tools used. We used protégé to create the ontology.

The last section of this paper about querying the ontology gives an idea of its possible uses. Appropriate user interfaces can be created to query the ontology which will lead to useful application such as nutrient calculator. Besides, this ontology can be extended by incorporating more facts about other vitamins and minerals. Number of such ontologies can be combined to create a knowledge base to be used in health care or other related fields.

7. REFERENCES

- [1] Noy, N., McGuinness, D.L, Ontology Development 101: A guide to creating your first ontology, Stanford Medical Informatics Technical Report No SMI-2001-0880, 2000
- [2] <http://protege.stanford.edu/>
- [3] <http://ods.od.nih.gov/factsheets/Vitamina-HealthProfessional/>
- [4] <http://www.who.int/vmnis/vitamina/prevallence/en/index.html>
- [5] Daconta, C., Obrst J. and Smith, T., The Semantic Web A Guide To The Future Of Xml, Web Services, And Knowledge Management, Wiley, Indianapolis, Indiana, 2003
- [6] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web, Scientific American, May 2001. Also <http://www.w3.org/2001/sw/>
- [7] Malik, S.K, Prakash, N. and Rizvi, S.A.M, Developing an University Ontology in Education Domain using Protégé for Semantic Web, *IJEST* Vol. 2(9), 2010
- [8] <http://www.vitamin-basics.com/>
- [9] Fernandez-Lopez, A Gomez-Perez, N Juristo. 1997. Methontology: from ontological art towards ontological engineering, Proceedings of the AAAI97 Spring, 1997.
- [10] Andrea Calí, Georg Gottlob, Andreas Pieris, "Ontological query answering under expressive Entity-Relationship schemata", *Journal Information Systems* Volume 37 Issue 4, June, 2012 Pages 320-335
- [11] Anton Froeyman," The Ontology of Causal Process Theories", *philosophia* September 2012, volume 4, issue 3, pages 523-538
- [12] Ken Kaneiwa, Philip H. P. Nguyen, "Sorted hyper-predicate knowledge bases for ontologies and rules", *Proceeding SAC '12 Proceedings of the 27th Annual ACM Symposium on Applied Computing* Pages 312-319